TiMAP: Laboratory for Text information, Mining, Analysis and Discovery

Mathew Palakal

School of Informatics

Indiana University – Purdue University Indianapolis

Abstract

The Laboratory for Text information Mining, Analysis, and Discovery, deal with mining associations, extracting knowledge, and perform network analysis from biomedical and health related literature documents. Currently we are mining literature documents associated with Colon Rectal Cancer (CRC), Limb Regeneration, Spices in Therapeutics, best practices in Geriatric Care, and Drug-Disease side effects in cancer treatments.

The work on CRC focuses on global analysis of bibliomic data, including literature, to identify novel biomarkers for Colon Rectal Cancer. In this work, we augment the experimental data of colorectal cancer with protein interaction data derived from literature to identify the biomarkers by analyzing all the entities, pathways, and networks at various stages of the disease propagation. In limb regeneration study, high throughput datasets are augmented with literature data to carry out a systems biology analysis to globally understand events during limb regeneration. Networks and pathways are used to visualize how different entities in these datasets are inter-related to carry out specific biological processes in a regenerating system. Under spices as therapeutics study, the goal is to harvest the scientific evidence on what spices are known to effective in treating various diseases and understanding how these spices really work at the molecular level. In this work, we are developing a knowledgebase called SpiceRDb, which is a unique attempt to elucidate the science behind the action of spices on various disease pathways, using text mining and molecular modeling tools. Geriatrics is the oen of the health related projects under TiMAP. Geriatrics is the study of caring for the elderly. We are working on building knowledgebase for a decision support system for different interventions in geriatrics. There exists a lot of literature about best practices in geriatric care. Our goal in this project is to discover best practice in geriatric as casual associations that are present in the literature. The challenge in mining literature for geriatrics is that there are no domain specific dictionaries. Identifying terms, building a dictionary, and defining relationship between the entities are the focus of this research. Finally, OncoMiner is a user-centric text information retrieval and knowledge discovery tool for cancer-related literature and navigation of the extracted knowledge as associations using a novel interactive information visualization tool. One such knowledge that is provided is the drug-side effect associations in chemotherapies.

In the TiMAP lab, various text mining and network analysis tools have been developed and we have demonstrated through these various projects that, literature knowledge is a powerful way to augment experimental data to discover and connect the “missing dots”. The dots are the unknown and relevant knowledge nuggets.